

Understanding the Hastings Algorithm

David D. L. Minh
Illinois Institute of Technology
BCHS Chemistry Division
3101 South Dearborn St.
Chicago, IL 60616

Do Le (Paul) Minh
Department of ISDS, California State University, Fullerton
CA 92831, USA
dminh@fullerton.edu

Abstract

The Hastings algorithm is a key tool in computational science. While mathematically justified by detailed balance, it can be conceptually difficult to grasp. Here, we present two complementary and intuitive ways to derive and understand the algorithm. In our framework, it is straightforward to see that the celebrated Metropolis-Hastings algorithm has the highest acceptance probability of all Hastings algorithms.

Keywords: Hastings algorithm; Metropolis-Hastings algorithm; Markov chain Monte Carlo; Simulation.

Mathematics Subject Classification: Primary 65C05; Secondary 78M31, 80M31.

1 Introduction

1.1 The Hastings algorithm (HA)

The Hastings algorithm (HA) (Hastings, 1970) is a stochastic sampling technique widely used throughout computational science. As a Markov Chain Monte Carlo method, HA does not attempt to generate a sequence of independent samples from a “target distribution” $\pi(\cdot)$, defined on the state space (E, \mathcal{E}) , but rather a Markov chain $\{X_n, n = 1, 2, 3, \dots\}$ having $\pi(\cdot)$ as its invariant distribution. Although variates in the chain are not independent, they may nonetheless be used to estimate statistical expectations with respect to $\pi(\cdot)$. (In a slight abuse of notation, we will often use the same symbol to denote both a measure and its density function.)

In many applications, the target distribution takes the form $\pi(\cdot) = p(\cdot)/P$, where the normalizing constant $P = \int_E p(x) dx$ is unknown. We call $p(\cdot)$ the un-normalized target distribution and $\pi(\cdot)$ the normalized one. If x is a variate generated from $\pi(\cdot)$, we may interchangeably write $x \sim \pi(\cdot)$ or $x \sim p(\cdot)$.

Let $U(0, 1)$ represent the uniform distribution on $(0, 1)$. In order to use all subsequently described algorithms, given $X_n = x$, we require a “proposal density” $\gamma(\cdot|x)$ which may (or may not) depend on x , and whose variates can be generated by other means.

Given $X_n = x \sim \pi(\cdot)$, we can generate $X_{n+1} \sim \pi(\cdot)$ by

Algorithm HA (Hastings)

HA1. generate $y \sim \gamma(\cdot|x)$ and $r \sim U(0, 1)$

HA2. if $r \leq \alpha_{HA}(x, y)$, output $X_{n+1} = y$

HA3. else, output $X_{n+1} = x$

where $\alpha_{HA}(x, y)$ is the Hastings’ “acceptance probability,” defined in terms of a symmetric function $s(\cdot, \cdot)$ that satisfies the following condition: For all $x, y \in E$,

$$0 \leq \alpha_{HA}(x, y) = s(x, y) \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} \leq 1. \quad (1)$$

(In Equation (6) in Hastings, 1970, this condition was expressed in terms of the normalized $\pi(\cdot)$, rather than the un-normalized $p(\cdot)$.)

1.2 Some special forms of the Hastings algorithm

1.2.1 The Metropolis-Hastings algorithm (MH)

HA was introduced as a generalization of the previously known Metropolis (1953) and Barker (1965) algorithms. In the celebrated paper by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953), the proposal densities are assumed to be symmetric (that is, $\gamma(x|y) = \gamma(y|x)$) and the acceptance probability in Step HA2 is,

$$\alpha_{MT}(x, y) = \min \left\{ \frac{p(y)}{p(x)}, 1 \right\}.$$

Hastings generalized the Metropolis algorithm into the well-known Metropolis-Hastings algorithm (MH) by setting $s(x, y) = s_{MH}(x, y)$, where,

$$\begin{aligned} s_{MH}(x, y) &= \begin{cases} 1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} & \text{if } \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)} \geq 1 \\ 1 + \frac{p(y)}{\gamma(y|x)} \frac{\gamma(x|y)}{p(x)} & \text{if } \frac{\gamma(y|x)}{p(y)} \frac{p(x)}{\gamma(x|y)} \geq 1 \end{cases} \\ &= \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\}. \end{aligned} \quad (2)$$

The acceptance probability $\alpha_{HA}(x, y)$ in Equation (1) then becomes the well-known MH acceptance probability (Chib and Greenberg, 1995 and Tierney, 1994):

$$\alpha_{MH}(x, y) = \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\}. \quad (3)$$

1.2.2 The Barker algorithm (BK)

Barker (1965) proposed the following acceptance probability, which uses the symmetric proposal densities $\gamma(x|y) = \gamma(y|x)$,

$$\alpha_{BK}^{(s)}(x, y) = \left(1 + \frac{p(x)}{p(y)} \right)^{-1},$$

which Hastings generalized by setting $s(x, y) = 1$ in Equation (1):

$$\alpha_{BK}(x, y) = \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1}. \quad (4)$$

We will subsequently refer to the Hastings algorithm with the acceptance probability α_{BK} as the Barker algorithm (BK).

1.2.3 Another special form of HA

As another example of HA, consider the case where $s(x, y)$ takes the following symmetric form:

$$\begin{aligned} s(x, y) &= \min\left(\frac{\gamma(x|y)}{p(x)}, 1\right) \min\left(\frac{\gamma(y|x)}{p(y)}, 1\right) \left(\frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)}\right) \\ &= \min\left(\frac{\gamma(x|y)}{p(x)}, 1\right) \min\left(\frac{p(y)}{\gamma(y|x)}, 1\right) \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)}\right). \end{aligned} \quad (5)$$

Substituting this form of $s(x, y)$ into Equation (1) results in the following acceptance probability for all $x, y \in E$:

$$\min\left(\frac{\gamma(x|y)}{p(x)}, 1\right) \min\left(\frac{p(y)}{\gamma(y|x)}, 1\right) \leq 1. \quad (6)$$

1.3 The detailed balance

To prove that a Markov chain $\{X_n, n = 1, 2, 3, \dots\}$ has the invariant distribution $\pi(\cdot)$, it is sufficient to show that its transition kernel $P(\cdot|\cdot)$ satisfies detailed balance (which is also called the “reversibility condition”) with respect to $p(\cdot) = P\pi(\cdot)$; that is, for all $x, y \in E$,

$$p(x)P(y|x) = P(x|y)p(y).$$

In this paper, all transition kernels can be expressed in two parts,

$$P(y|x) = r_1(y|x) + I(x = y)r_2(y|x),$$

where $I(a) = 1$ if a is true, 0 otherwise. Because $p(x)I(x = y)r_2(y|x) = p(y)I(x = y)r_2(x|y)$, for notational simplicity, we only prove detailed balance for $x \neq y$, omitting the second part.

For HA, the transition kernel for all $x, y \in E$ is,

$$P_{HA}(y|x) = \alpha_{HA}(x, y)\gamma(y|x) + I(x = y) \left[\int_E (1 - \alpha_{HA}(x, z))\gamma(z|x) dz \right]. \quad (7)$$

The first term is the probability of the proposed variate $y \sim \gamma(\cdot|x)$ being accepted (the chain moves to y). The term inside the integration is probability of the proposed variate $z \sim \gamma(\cdot|x)$ being rejected (the chain remains at

x). Thus $P_{HA}(y|x)$ satisfies detailed balance with respect to $p(\cdot)$ because, from Equation (1), for all $x, y \in E$ and $x \neq y$,

$$\begin{aligned} p(x)P_{HA}(y|x) &= p(x)\alpha_{HA}(x, y)\gamma(y|x) \\ &= p(x)s(x, y)\frac{p(y)\gamma(x|y)}{p(x)\gamma(y|x) + p(y)\gamma(x|y)}\gamma(y|x) \\ &= P_{HA}(x|y)p(y). \end{aligned}$$

While verifying that HA satisfies detailed balance is simple, conceptually understanding it is much harder. In a paper interpreting MH geometrically, Billera & Diaconis (2001) wrote, “The algorithm is widely used for simulations in physics, chemistry, biology and statistics. It appears as the first entry of a recent list of great algorithms of 20th-century scientific computing [4]. Yet for many people (including the present authors) the Metropolis-Hastings algorithm seems like a magic trick. It is hard to see where it comes from or why it works.” (Reference [4] refers to Dongarra and Sullivan, 2000.) If it is hard to conceptually understand the development of MH, it is even harder to visualize the more general HA.

In this paper, we provide two complementary and intuitive derivations of the Hastings algorithm. First, we present a new form of the acceptance probability in the next section.

2 Algorithm M

2.1 Algorithm M

Given $X_n = x \sim \pi(\cdot)$, $X_{n+1} \sim \pi(\cdot)$ can be generated by

Algorithm M

- M1. generate $y \sim \gamma(\cdot|x)$ and $r \sim U(0, 1)$
- M2. if $r \leq \alpha_M(x, y)$, output $X_{n+1} = y$
- M3. else, output $X_{n+1} = x$

in which the acceptance probability $\alpha_M(x, y)$ is, for all $x, y \in E$,

$$\alpha_M(x, y) = \min \left\{ \frac{k(x, y)\gamma(x|y)}{p(x)}, 1 \right\} \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} \leq 1, \quad (8)$$

where $k(\cdot, \cdot) : E \times E \rightarrow R > 0$ is a *symmetric* function.

Similar to Equation (7), the transition kernel of Algorithm M is, for all $x, y \in E$,

$$P_M(y|x) = \alpha_M(x, y)\gamma(y|x) + I(x = y) \left[\int_E (1 - \alpha_M(x, z))\gamma(z|x)dz \right].$$

$P_M(y|x)$ satisfies detailed balance with respect to $p(\cdot)$: For all $x, y \in E$ and $x \neq y$,

$$\begin{aligned} p(x)P_M(y|x) &= p(x)\alpha_M(x, y)\gamma(y|x) \\ &= p(x) \min \left\{ \frac{k(x, y)\gamma(x|y)}{p(x)}, 1 \right\} \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} \gamma(y|x) \\ &= \min \{k(x, y)\gamma(x|y), p(x)\} \min \{p(y), k(x, y)\gamma(y|x)\} \frac{1}{k(x, y)} \\ &= P_M(x|y)p(y). \end{aligned}$$

If $k(x, y)$ is a positive constant k , then $p(\cdot)/k$ is just another un-normalized distribution corresponding to $\pi(\cdot)$ and the acceptance probability $\alpha_M(x, y)$ in Equation (8) is the same as that in Equation (6). So, for the rest of this paper, we exclude the case in which $k(x, y) = k$.

For all $x, y \in E$, we define,

$$L(x, y) = \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} \text{ and } H(x, y) = \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\}.$$

When $L(x, y) < k(x, y) < H(x, y)$, from Equation (8),

$$\begin{aligned} \alpha_M(x, y) &= \begin{cases} 1 & \text{if } \frac{p(x)}{\gamma(x|y)} < k(x, y) < \frac{p(y)}{\gamma(y|x)} \\ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)} & \text{if } \frac{p(y)}{\gamma(y|x)} < k(x, y) < \frac{p(x)}{\gamma(x|y)} \end{cases} \\ &= \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\}. \end{aligned} \quad (9)$$

Thus the acceptance probability $\alpha_M(x, y)$ may be expressed as a piecewise function that depends on the relationship between $k(x, y)$, $L(x, y)$, and $H(x, y)$:

$$\alpha_M(x, y) = \begin{cases} \frac{p(y)}{k(x, y)\gamma(y|x)} & \text{if } k(x, y) \geq H(x, y) \\ \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\} & \text{if } L(x, y) < k(x, y) < H(x, y) \\ \frac{k(x, y)\gamma(x|y)}{p(x)} & \text{if } k(x, y) \leq L(x, y) \end{cases} \quad (10)$$

From Equations (3) and (9), it is clear that MH is a special case of Algorithm M when $L(x, y) < k(x, y) < H(x, y)$.

BK, with acceptance probability $\alpha_{BK}(x, y)$ in Equation (4), can also be shown to be a special case of Algorithm M : We set,

$$k(x, y) = \frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)} \geq \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = H(x, y).$$

Then from Equation (10),

$$\alpha_M(x, y) = \frac{p(y)}{\gamma(y|x)} \left(\frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)} \right)^{-1} = \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \alpha_{BK}(x, y).$$

We can also set,

$$\begin{aligned} k(x, y) &= \left(\frac{\gamma(x|y)}{p(x)} + \frac{\gamma(y|x)}{p(y)} \right)^{-1} \\ &\leq \left(\max \left\{ \frac{\gamma(x|y)}{p(x)}, \frac{\gamma(y|x)}{p(y)} \right\} \right)^{-1} = \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = L(x, y), \end{aligned}$$

to obtain the same BK acceptance probability from Equation (10):

$$\alpha_M(x, y) = \frac{\gamma(x|y)}{p(x)} \left(\frac{\gamma(x|y)}{p(x)} + \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \alpha_{BK}(x, y).$$

2.2 Algorithm M and HA

We now show that HA and Algorithm M are equivalent. First, we show that the former is a special case of the latter. We then show that the latter is a special case of the former.

2.2.1 HA is a special case of Algorithm M

HA is a special case of Algorithm M if, for any acceptance probability $\alpha_{HA}(\cdot, \cdot)$ in HA, expressed in terms of $s(\cdot, \cdot)$, we can find the same acceptance probability $\alpha_M(\cdot, \cdot)$ in Algorithm M : For each $s(\cdot, \cdot)$ satisfying the Hastings condition (1), we define the following symmetric function,

$$M_s(x, y) = \frac{1}{s(x, y)} \left(\frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)} \right) = \frac{1}{s(x, y)} \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \frac{p(y)}{\gamma(y|x)} \geq \frac{p(y)}{\gamma(y|x)}. \quad (11)$$

Because $M_s(x, y)$ is symmetric, we also have $M_s(x, y) \geq p(x)/\gamma(x|y)$; hence $M_s(x, y) \geq H(x, y)$. Now letting $k(x, y) = M_s(x, y)$ in Equation (10), we find in Algorithm M the same acceptance probability as that defined by $s(x, y)$ in HA:

$$\alpha_M(x, y) = \frac{p(y)}{M_s(x, y)\gamma(y|x)} = s(x, y) \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \alpha_{HA}(x, y). \quad (12)$$

For example, if $s(x, y)$ takes the form of Equation (5), then Equation (11) yields

$$\begin{aligned} M_s(x, y) &= \left\{ \min \left(\frac{\gamma(x|y)}{p(x)}, 1 \right) \min \left(\frac{\gamma(y|x)}{p(y)}, 1 \right) \right\}^{-1} \\ &\geq \left\{ \min \left(\frac{\gamma(x|y)}{p(x)}, \frac{\gamma(y|x)}{p(y)} \right) \right\}^{-1} = \max \left(\frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right) = H(x, y). \end{aligned}$$

Set $k(x, y) = M_s(x, y)$, Algorithm M yields the same acceptance probability as the special form of $\alpha_{HA}(x, y)$ in Equation (6):

$$\begin{aligned} \alpha_M(x, y) &= \frac{p(y)}{M_s(x, y)\gamma(y|x)} = \min \left(\frac{\gamma(x|y)}{p(x)}, 1 \right) \min \left(\frac{\gamma(y|x)}{p(y)}, 1 \right) \frac{p(y)}{\gamma(y|x)} \\ &= \min \left(\frac{\gamma(x|y)}{p(x)}, 1 \right) \min \left(\frac{p(y)}{\gamma(y|x)}, 1 \right). \end{aligned}$$

Return to the general $s(\cdot, \cdot)$ satisfying the Hastings condition (1), we may also define the following symmetric function,

$$m_s(x, y) = s(x, y) \left(\frac{\gamma(x|y)}{p(x)} + \frac{\gamma(y|x)}{p(y)} \right)^{-1} = s(x, y) \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} \frac{p(x)}{\gamma(x|y)} \leq \frac{p(x)}{\gamma(x|y)}. \quad (13)$$

Because of its symmetrical property, $m_s(x, y) \leq L(x, y)$. With $k(x, y) = m_s(x, y)$ in Equation (10), we obtain in Algorithm M the same acceptance probability as that defined by $s(x, y)$ in HA:

$$\alpha_M(x, y) = \frac{m_s(x, y)\gamma(x|y)}{p(x)} = s(x, y) \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \alpha_{HA}(x, y). \quad (14)$$

For example, if $s(x, y)$ takes the form of Equation (5), then Equation (13) yields

$$\begin{aligned} m_s(x, y) &= \min \left(\frac{\gamma(x|y)}{p(x)}, 1 \right) \min \left(\frac{p(y)}{\gamma(y|x)}, 1 \right) \frac{p(x)}{\gamma(x|y)} \\ &\leq \min \left(\frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right) = L(x, y). \end{aligned}$$

When we set $k(x, y) = m_s(x, y)$, Algorithm M yields the same acceptance probability as the special form of $\alpha_{HA}(x, y)$ in Equation (6):

$$\alpha_M(x, y) = \frac{m_s(x, y)\gamma(x|y)}{p(x)} = \min \left(\frac{\gamma(x|y)}{p(x)}, 1 \right) \min \left(\frac{p(y)}{\gamma(y|x)}, 1 \right).$$

In the next subsection the reverse is proven.

2.2.2 Algorithm M is a special case of HA

Algorithm M is a special case of HA if, for any acceptance probability $\alpha_M(\cdot, \cdot)$ in Algorithm M (expressed in terms of $k(\cdot, \cdot)$), we can find the same acceptance probability $\alpha_{HA}(\cdot, \cdot)$ in HA.

Case 1: When $k(x, y) \geq H(x, y)$: We set,

$$\begin{aligned} s(x, y) &= \frac{1}{k(x, y)} \left(\frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)} \right) = \frac{p(y)}{k(x, y)\gamma(y|x)} \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \\ &\leq 1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)}. \end{aligned}$$

Substituting this form of $s(x, y)$ into Equation (1) we obtain in HA the same acceptance probability as $\alpha_M(x, y)$ in Equation (10) when $k(x, y) \geq H(x, y)$.

Case 2: When $L(x, y) < k(x, y) < H(x, y)$: Equation (9) gives $\alpha_M(x, y) = \alpha_{MH}(x, y)$. We thus set $s(x, y) = s_{MH}(x, y)$, as defined in Equation (2), to obtain the same acceptance probability.

Case 3: When $k(x, y) \leq L(x, y)$: We define,

$$\begin{aligned} s(x, y) &= k(x, y) \left(\frac{\gamma(x|y)}{p(x)} + \frac{\gamma(y|x)}{p(y)} \right) = \frac{k(x, y)\gamma(x|y)}{p(x)} \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \\ &\leq 1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)}. \end{aligned}$$

Substituting this form of $s(x, y)$ into Equation (1) we obtain in HA the same acceptance probability as $\alpha_M(x, y)$ in Equation (10) when $k(x, y) \leq L(x, y)$.

Because Algorithm M is a special case of HA and HA is also a special case of Algorithm M , they are equivalent. It is worth noting, however, that the relationship between $s(\cdot, \cdot)$ and $k(\cdot, \cdot)$ is not one-to-one. The set of all $k(x, y) > 0$ available to construct $\alpha_M(x, y)$ is larger than the set of all $s(x, y) > 0$ available to construct $\alpha_{HA}(x, y)$, because $s(x, y)$ must also satisfy the Hastings' condition in Equation (1). In fact, for every $s(x, y)$, there are at least two distinct expressions for $k(x, y)$: $M_s(x, y) \geq H(x, y)$ as defined in Equation (11), and $m_s(x, y) \leq L(x, y)$, as defined in Equation (13). As shown in Equation (9), all functions $k(x, y)$ that satisfy $L(x, y) < k(x, y) < H(x, y)$ may be mapped to $s_{MH}(x, y)$.

2.3 Algorithm M and the Stein Algorithm

Stein (in Liu 2001, p. 112) proposed an algorithm similar to HA in which the acceptance probability $\alpha_{ST}(x, y)$ is expressed in terms of a symmetric function $\delta(\cdot, \cdot)$ such that,

$$0 \leq \alpha_{ST}(x, y) = \frac{\delta(x, y)}{p(x)\gamma(y|x)} \leq 1. \quad (15)$$

By the same logic with which we showed the equivalence of Algorithm M and HA, we can show the equivalence of Algorithm M and the Stein algorithm.

2.3.1 The Stein algorithm is a special case of Algorithm M

For each acceptance probability $\alpha_{ST}(\cdot, \cdot)$, expressed in terms of $\delta(\cdot, \cdot)$, we can find the same acceptance probability $\alpha_M(\cdot, \cdot)$: We define the symmetric function,

$$M_\delta(x, y) = \frac{p(x)p(y)}{\delta(x, y)} \geq \frac{p(x)p(y)}{p(x)\gamma(y|x)} = \frac{p(y)}{\gamma(y|x)}.$$

By symmetry, $M_\delta(x, y) \geq H(x, y)$. Then with $k(x, y) = M_\delta(x, y)$ in Equation (10) we obtain in Algorithm M the same acceptance probability as that defined by $\delta(x, y)$ in the Stein algorithm:

$$\alpha_M(x, y) = \frac{\delta(x, y)}{p(x)p(y)} \frac{p(y)}{\gamma(y|x)} = \frac{\delta(x, y)}{p(x)\gamma(y|x)} = \alpha_{ST}(x, y).$$

Alternatively, we can define the symmetric function,

$$m_\delta(x, y) = \frac{\delta(x, y)}{\gamma(x|y)\gamma(y|x)} \leq \frac{p(x)\gamma(y|x)}{\gamma(x|y)\gamma(y|x)} = \frac{p(x)}{\gamma(x|y)}.$$

By symmetry, $m_\delta(x, y) \leq L(x, y)$. Then with $k(x, y) = m_\delta(x, y)$ in Equation (10) we obtain in Algorithm M the same acceptance probability as that defined by $s(x, y)$ in the Stein algorithm:

$$\alpha_M(x, y) = \frac{\delta(x, y)}{\gamma(x|y)\gamma(y|x)} \frac{\gamma(x|y)}{p(x)} = \frac{\delta(x, y)}{p(x)\gamma(y|x)} = \alpha_{ST}(x, y).$$

2.3.2 Algorithm M is a special case of the Stein algorithm

For each acceptance probability $\alpha_M(\cdot, \cdot)$ expressed in terms of $k(\cdot, \cdot)$, we can find the same acceptance probability $\alpha_{ST}(\cdot, \cdot)$:

Case 1: When $k(x, y) \geq H(x, y)$: We set

$$\begin{aligned} \delta(x, y) &= \frac{p(x)p(y)}{k(x, y)} \leq p(x)p(y) \left(\max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} \right)^{-1} \\ &= p(x)p(y) \min \left\{ \frac{\gamma(x|y)}{p(x)}, \frac{\gamma(y|x)}{p(y)} \right\} \\ &= \min \{ p(y)\gamma(x|y), p(x)\gamma(y|x) \} \leq p(x)\gamma(y|x). \end{aligned}$$

Substituting this form of $\delta(x, y)$ into Equation (15) we obtain in the Stein algorithm the same acceptance probability as $\alpha_M(x, y)$ in Equation (10) when $k(x, y) \geq H(x, y)$.

Case 2: When $L(x, y) < k(x, y) < H(x, y)$: Due to Equation (9), we set

$$\delta(x, y) = \min \{p(y)\gamma(x|y), p(x)\gamma(y|x)\} \leq p(x)\gamma(y|x),$$

to obtain the same acceptance probability:

$$\alpha_{ST}(x, y) = \frac{\min \{p(y)\gamma(x|y), p(x)\gamma(y|x)\}}{p(x)\gamma(y|x)} = \alpha_{MH}(x, y).$$

Case 3: When $k(x, y) \leq L(x, y)$: We set

$$\begin{aligned} \delta(x, y) &= k(x, y)\gamma(x|y)\gamma(y|x) \leq \gamma(x|y)\gamma(y|x) \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} \\ &= \min \{p(x)\gamma(y|x), p(y)\gamma(x|y)\} \leq p(x)\gamma(y|x) \end{aligned}$$

Substituting this form of $\delta(x, y)$ into Equation (15) we obtain in the Stein algorithm the same acceptance probability as $\alpha_M(x, y)$ in Equation (10) when $k(x, y) \leq L(x, y)$.

Previously, the relationship between the Stein algorithm and HA was unclear. Now we have shown that Algorithm M , the Stein algorithm and HA are all equivalent.

If Algorithm M is equivalent to HA, then why do we introduce it? For the rest of this paper, we will show how Algorithm M may be developed intuitively; we do not merely have to accept it because it satisfies detailed balance. In the following section, we describe how Algorithm M may be obtained from a series of incremental modifications to the Acceptance-Rejection (AR) algorithm (von Neumann, 1951).

3 Markovian Acceptance-Rejection (MAR)

3.1 Acceptance-Rejection (AR)

AR is a well-known algorithm that uses a proposal density $\gamma(\cdot)$ to generate a sequence of independent variates from $p(\cdot)$. It requires a “majorizing coefficient” M such that the “majorizing function” $M\gamma(\cdot)$ satisfies $M\gamma(z) \geq p(z)$ for all $z \in E$. Given $X_n = x \sim \pi(\cdot)$, then $X_{n+1} \sim \pi(\cdot)$ can be generated by

Algorithm AR (Acceptance-Rejection)

AR1. set $reject = 1$

AR2. while $reject = 1$

AR2a. generate $y \sim \gamma(\cdot)$ and $r \sim U(0, 1)$

AR2b. if $r \leq \frac{p(y)}{M\gamma(y)} \leq 1$, output $X_{n+1} = y$, set $reject = 0$

AR3. endwhile

AR is easy to understand conceptually. The condition $M\gamma(\cdot) \geq p(\cdot)$ assures that the surface $M\gamma(\cdot)$ is above that of $p(\cdot)$. With $r \sim U(0, 1)$, every pair $(y \sim \gamma(\cdot), rM\gamma(y))$ is uniformly distributed under the surface $M\gamma(\cdot)$. Of these, those that satisfy the condition in Step AR2b (and hence are accepted) are also uniformly distributed under the surface $p(\cdot)$. These y variates have density $\pi(\cdot)$. (See Minh, 2001, Chap. 13.)

3.2 Independence Markovian Acceptance-Rejection (IMAR)

We now present a simple modification of AR into what we call the “Independence Markovian Acceptance-Rejection” algorithm (IMAR). Given $X_n = x \sim \pi(\cdot)$, and a proposed density $\gamma(\cdot)$ (which is independent of x), $X_{n+1} \sim \pi(\cdot)$ can be generated by

Algorithm IMAR (Independence Markovian Acceptance-Rejection)

IMA1. generate $y \sim \gamma(\cdot)$ and $r \sim U(0, 1)$

IMA2. if $r \leq \alpha_{IMA}(x, y) = \frac{p(y)}{M\gamma(y)} \leq 1$, output $X_{n+1} = y$

IMA3. else, output $X_{n+1} = x$

The main distinction between AR and IMAR is that, when a proposed variate y is rejected in IMAR, the variate x is repeated. Suppose that a sequence of proposed variates is $y_1, y_2, y_3, y_4, y_5, y_6, \dots$. If AR accepts y_1, y_3, y_6, \dots , then with the same sequence of random numbers, IMAR would generate $y_1, y_1, y_3, y_3, y_3, y_6, \dots$. Because of the repetitions, IMAR does not generate

independent variates; rather it is a Markov Chain Monte Carlo method that satisfies detailed balance with respect to $p(\cdot)$: For all $x, y \in E$ and $x \neq y$,

$$p(x)P_{IMA}(y|x) = p(x)\frac{p(y)}{M\gamma(y)}\gamma(y) = P_{IMA}(x|y)p(y).$$

As in AR, the expected number of times that z is delivered in Step IMA2 in a simulation is proportional to $p(z)$. Also, the expected number of duplications in Step IMA3 is the same for all variates, which is $M - 1$, the expected number of consecutive rejections in the corresponding AR. Thus the expected total number of times that z and its duplicates are delivered is proportional to $Mp(z)$, or to $\pi(z)$ because M is a constant.

While it is hard to find a majorizing coefficient M such that $M\gamma(z) \geq p(z)$ for all $z \in E$, it is easier to find a “deficient” majorizing coefficient M such that $M\gamma(z) \geq p(z)$ for some $z \in E$. In this case, it is well known that AR produces variates from $\min\{p(\cdot), M\gamma(\cdot)\}$. This is also true for IMAR, in which $M\gamma(\cdot)$ serves as the majorizing function for $\min\{p(\cdot), M\gamma(\cdot)\}$, resulting in the acceptance probability

$$\alpha_D(x, y) = \frac{\min\{p(y), M\gamma(y)\}}{M\gamma(y)} = \min\left\{\frac{p(y)}{M\gamma(y)}, 1\right\}.$$

For future reference, it is important to note that, even with a deficient majorizing constant, AR and IMAR still generate variates $y \sim \pi(\cdot)$ within the region $\{z : M\gamma(z) \geq p(z)\}$.

3.3 Markovian Acceptance-Rejection (MAR)

As a generalization of IMAR, we now allow the proposal density $\gamma(\cdot|x)$ to be dependent on the chain’s current value $X_n = x$.

If we knew beforehand that AR accepts y_3 out of 3 proposed variates y_1, y_2 and y_3 , then all we would need is a majorizing coefficient M such that $M\gamma(y_i) \geq p(y_i)$ for $i = 1, 2, 3$. The problem is that the number of consecutive rejections before an acceptance in AR may be infinite, and y can be anywhere in E . Furthermore, to generate independent variates, the majorizing coefficient M in AR must be independent of the current variate x . So AR needs an “absolute” majorizing coefficient M such that $M\gamma(z) \geq p(z)$ for all $z \in E$.

When the proposal density $\gamma(\cdot|x)$ is allowed to be dependent on $X_n = x$, the requirement of an absolute majorizing coefficient M , however, is too

restrictive: if there is a pair $\eta, \xi \in E$ such that $\gamma(\eta|\xi) = 0$ and $p(\eta) > 0$, then we must have $M = \infty$.

Fortunately, similar to IMAR, in the following Algorithm MAR, which allows $\gamma(\cdot|x)$ to be dependent on x , either the current variate x or the proposed variate y must be delivered in each iteration. So, instead of requiring an absolute majorizing coefficient M , we only need a “relative” majorizing coefficient $M(\cdot, \cdot) > 0$ that may change with each pair (x, y) , so long as, for all $x, y \in E$, $M(x, y)\gamma(x|y) \geq p(x)$ and $M(x, y)\gamma(y|x) \geq p(y)$, or,

$$M(x, y) \geq \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = H(x, y). \quad (16)$$

It is necessary that the relative majorizing coefficient $M(\cdot, \cdot)$ is symmetric, in order to preserve the balance of flows from x to y and from y to x . We thus may write $M(\cdot, \cdot)$ in terms of any symmetric function $C(\cdot, \cdot) \geq 1$:

$$M(x, y) = C(x, y) \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\}. \quad (17)$$

Given $X_n = x \sim \pi(\cdot)$, the following algorithm may be used to generate $X_{n+1} \sim \pi(\cdot)$:

Algorithm MAR (Markovian Acceptance-Rejection)

MA1. generate $y \sim \gamma(\cdot|x)$ and $r \sim U(0, 1)$

MA2. if $r \leq \alpha_{MA}(x, y)$, output $X_{n+1} = y$

MA3. else, output $X_{n+1} = x$

where, with $M(x, y)$ defined in Equation (17),

$$\alpha_{MA}(x, y) = \frac{p(y)}{M(x, y)\gamma(y|x)} \quad (18)$$

$$= \frac{p(y)}{C(x, y) \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} \gamma(y|x)} \quad (19)$$

$$= \frac{1}{C(x, y)} \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\} \leq 1. \quad (20)$$

As with IMAR, it is straightforward to show that, if $M(\cdot, \cdot)$ is symmetric, then the transition kernel MAR satisfies detailed balance with respect to $p(\cdot)$: For all $x, y \in E$ and $x \neq y$,

$$p(x)P_{MA}(y|x) = p(x)\frac{p(y)}{M(x, y)\gamma(y|x)}\gamma(y|x) = P_{MA}(x|y)p(y).$$

As previously noted, using a deficient (absolute) majorizing coefficient M in IMAR still generates variates $y \sim p(\cdot)$ within the region $\{z : M\gamma(z) \geq p(z)\}$. A relative majorizing coefficient may be a deficient (absolute) majorizing coefficient, but is sufficient for x and y , because both x and y are within the region $\{z : M(x, z)\gamma(z|x) \geq p(z)\}$.

We now show that BK and MH are two special cases of MAR.

3.4 BK in MAR

As a special case of MAR, we let

$$C(x, y) = \left(\frac{p(y)}{\gamma(y|x)} + \frac{p(x)}{\gamma(x|y)} \right) \left(\max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} \right)^{-1} > 1. \quad (21)$$

Then the acceptance probability in Equation (19) becomes the Barker's acceptance probability $\alpha_{BK}(x, y)$ in Equation (4):

$$\alpha_{MA}(x, y) = \frac{p(y)}{\left(\frac{p(y)}{\gamma(y|x)} + \frac{p(x)}{\gamma(x|y)} \right) \gamma(y|x)} = \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right)^{-1} = \alpha_{BK}(x, y).$$

3.5 MH in MAR

If we set $C(x, y) = 1$, the acceptance probability $\alpha_{MA}(x, y)$ in Equation (20) become the acceptance probability $\alpha_{MH}(x, y)$ in Equation (3) and MAR becomes MH.

Peskun (1973) introduced partial ordering on transition kernels to prove that, with the same proposal densities, $\alpha_{MH}(x, y)$ is optimal in terms of minimizing the asymptotic variance of sample path averages. In the MAR framework, it is straightforward to see that the acceptance probability in

Equation (20) is maximized when $C(x, y) = 1$. Thus MH has the highest acceptance probability of all Hastings algorithms.

Conceptually, the most efficient majorizing function $M\gamma(\cdot)$ in AR is the one that “touches” the target density $p(\cdot)$ at one point. Similarly, when $C(x, y) = 1$, Equation (17) shows that either $M(x, y)\gamma(x|y) = p(x)$ or $M(x, y)\gamma(y|x) = p(y)$. Any higher value of $C(x, y)$ only results in unnecessarily rejecting some proposed variates. This is what happens in BK, where $C_{BK}(x, y) > 1$ as in Equation (21).

We have derived and explained MAR intuitively. It turns out that MAR is equivalent to Algorithm M .

3.6 MAR and Algorithm M

MAR is a special case of Algorithm M if, for any acceptance probability $\alpha_{MA}(\cdot, \cdot)$ (defined in terms of the relative majorizing coefficient $M(x, y)$) in MAR, we can find the same acceptance probability $\alpha_M(\cdot, \cdot)$ in Algorithm M . We achieve this simply by letting $k(x, y) = M(x, y) \geq H(x, y)$ in Equation (10), resulting in $\alpha_M(x, y) = \alpha_{MA}(x, y)$.

For equivalence, the reverse must also be true; that is, Algorithm M is a special case of MAR. We now show that, for any acceptance probability $\alpha_M(x, y)$ (defined in terms of $k(x, y)$), we can also find the same $\alpha_{MA}(x, y)$ in MAR. Consider

$$M_k(x, y) = k(x, y) \max \left\{ \frac{p(x)}{k(x, y)\gamma(x|y)}, 1 \right\} \max \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\}.$$

$M_k(x, y)$ is a relative majorizing coefficient because it satisfies the inequality (16):

Case 1. When $k(x, y) \geq H(x, y)$:

$$M_k(x, y) = k(x, y) \geq H(x, y)$$

Case 2. When $L(x, y) < k(x, y) < H(x, y)$:

$$\begin{aligned} M_k(x, y) &= \begin{cases} \frac{p(y)}{\gamma(y|x)} & \text{if } \frac{p(x)}{\gamma(x|y)} \leq k(x, y) \leq \frac{p(y)}{\gamma(y|x)} \\ \frac{p(x)}{\gamma(x|y)} & \text{if } \frac{p(y)}{\gamma(y|x)} \leq k(x, y) \leq \frac{p(x)}{\gamma(x|y)} \end{cases} \\ &= \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = H(x, y) \end{aligned} \quad (22)$$

Case 3. If $k(x, y) \leq L(x, y)$:

$$\begin{aligned} M_k(x, y) &= \frac{1}{k(x, y)} \frac{p(x)}{\gamma(x|y)} \frac{p(y)}{\gamma(y|x)} \\ &\geq \frac{1}{\min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\}} \frac{p(x)}{\gamma(x|y)} \frac{p(y)}{\gamma(y|x)} = \max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = H(x, y) \end{aligned} \quad (23)$$

Setting $M(x, y) = M_k(x, y)$ in Equation (18) yields:

$$\begin{aligned} \alpha_{MA}(x, y) &= \frac{p(y)}{k(x, y) \max \left\{ \frac{p(x)}{k(x, y)\gamma(x|y)}, 1 \right\} \max \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} \gamma(y|x)} \\ &= \min \left\{ \frac{k(x, y)\gamma(x|y)}{p(x)}, 1 \right\} \min \left\{ \frac{k(x, y)\gamma(y|x)}{p(y)}, 1 \right\} \frac{p(y)}{k(x, y)\gamma(y|x)} \\ &= \min \left\{ \frac{k(x, y)\gamma(x|y)}{p(x)}, 1 \right\} \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} = \alpha_M(x, y) \end{aligned}$$

Hence Algorithm M and MAR are equivalent.

3.7 MAR and HA

Because MAR is equivalent to Algorithm M , and Algorithm M is equivalent to HA, MAR and HA are equivalent. To show this directly, for every $\alpha_{MA}(x, y)$ defined by $M(x, y) \geq H(x, y)$ in MAR, we set,

$$\begin{aligned} s(x, y) &= \frac{1}{M(x, y)} \left(\frac{p(x)}{\gamma(x|y)} + \frac{p(y)}{\gamma(y|x)} \right) = \frac{p(y)}{M(x, y)\gamma(y|x)} \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \\ &\leq 1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)}. \end{aligned} \quad (24)$$

Substituting this form of $s(x, y)$ into Equation (1) we obtain in HA the same acceptance probability $\alpha_{MA}(x, y)$. Hence MAR is a special case of HA. On the other hand, for each $\alpha_{HA}(x, y)$ defined by $s(x, y)$, we set $M(x, y) = M_s(x, y) \geq H(x, y)$, where $M_s(x, y)$ was defined in Equation (11). Like Equation (12), Equation (18) then yields in MAR the same acceptance probability as $\alpha_{HA}(x, y)$ in HA. Thus, HA is a special case of MAR. Therefore, MAR and HA are equivalent.

Equations (11) and (24) show that there is a one-to-one mapping between the set of all symmetric functions $s(x, y)$ satisfying Hastings' condition, Equation (1), and the set of all symmetric functions $M(x, y)$ in the form of Equation (17). However, unlike the mysterious $s(\cdot, \cdot)$, $M(\cdot, \cdot)$ has a very intuitive interpretation of being a relative majorizing coefficient.

Thus far, we have intuitively derived HA as MAR, which is Algorithm M in which $k(\cdot, \cdot)$ is sufficiently large to be a relative majorizing coefficient. We now show that HA can also be explained in terms of an algorithm “dual” to MAR, which is Algorithm M with a sufficiently small coefficient $k(\cdot, \cdot)$.

4 Markovian Minorizing (MIR)

4.1 Independence Markovian Minorizing (IMIR)

We now return to the assumption that the proposal densities are independent of the chain's current variate, or $\gamma(\cdot|x) = \gamma(\cdot)$. We also assume that the support of $\gamma(\cdot)$ includes that of $p(\cdot)$ and there is an “absolute minorizing coefficient” m such that $m\gamma(z) \leq p(z)$ for all $z \in E$.

Consider the following algorithm that we call the “Independence Markovian Minorizing” algorithm (IMIR): Given $X_n = x \sim \pi(\cdot)$, then $X_{n+1} \sim \pi(\cdot)$ can be generated by

Algorithm IMIR (Independence Markovian Minorizing)

IMI1. generate $y \sim \gamma(\cdot)$ and $r \sim U(0, 1)$

IMI2. if $r \leq \alpha_{IMI}(x, y) = \frac{m\gamma(x)}{p(x)} \leq 1$, output $X_{n+1} = y$

IMI3. else, output $X_{n+1} = x$

The transition kernel of this algorithm is $P_{IMI}(y|x) = \alpha_{IMI}(x, y)\gamma(y)$ for all $x, y \in E$ and $x \neq y$, which satisfies detailed balance with respect to $p(\cdot)$:

$$p(x)P_{IMI}(y|x) = p(x)\frac{m\gamma(x)}{p(x)}\gamma(y) = P_{IMI}(x|y)p(y).$$

We may not have an absolute minorizing coefficient m , but only a “deficient” minorizing coefficient m such that $m\gamma(z) \leq p(z)$ for some $z \in E$.

Using $m\gamma(\cdot)$ as the minorizing function for $\max\{p(\cdot), m\gamma(\cdot)\}$ in Algorithm IMIR, the acceptance probability $\alpha_{IMI}(x, y)$ becomes

$$\alpha_d(x, y) = \frac{m\gamma(x)}{\max\{p(x), m\gamma(x)\}} = \min\left\{\frac{m\gamma(x)}{p(x)}, 1\right\},$$

and Algorithm IMIR generates variates from $\max\{p(\cdot), m\gamma(\cdot)\}$. Note that, similar to the discussion for IMAR, even with a deficient minorizing constant, IMIR still generates variates $y \sim \pi(\cdot)$ within the region $\{z : m\gamma(z) \leq p(z)\}$.

We wrote Algorithm IMIR in the form consistent with that of all other algorithms in this paper. However, we do not need to generate $y \sim \gamma(\cdot)$ in Step IMI1 if $r > \alpha_{IMI}(x, y)$ in Step IMI2. For a more intuitive understanding, Algorithm IMIR can also be written as,

Algorithm IMJ

IMJ1. generate $r \sim U(0, 1)$

IMJ2. if $r \leq \alpha_{IMI}(x, y) = \frac{m\gamma(x)}{p(x)} \leq 1$, generate $y \sim \gamma(\cdot)$, output $X_{n+1} = y$

IMJ3. else, output $X_{n+1} = x$

In a simulation, the expected number of times that $x = y \sim \gamma(\cdot)$ is delivered in Step IMJ2 is proportional to $\gamma(x)$. Furthermore, for each x so delivered, it is duplicated until the first success in a sequence of Bernoulli trials with success probability $m\gamma(x)/p(x)$; the expected number of its duplications is $p(x)/[m\gamma(x)] \geq 1$. Thus in a simulation, the expected total number of times that x is delivered is proportional to $\gamma(x) \{p(x)/[m\gamma(x)]\} = p(x)/m$, or to $p(x)$ because m is a constant.

In Minh et al (2012) we used the minorizing coefficient m to make any Markov Chain Monte Carlo method regenerative.

4.2 Markovian Minorizing (MIR)

If the proposal density is dependent on x , taking the form $\gamma(\cdot|x)$, the requirement of an “absolute” minorizing coefficient m such that $m\gamma(y|x) \leq p(y)$ for all $x, y \in E$ is too restrictive, and often can only be satisfied when $m = 0$.

Fortunately, similar to MAR, given the current variate x and proposed variate y , there is no need for such an absolute minorizing coefficient, but only a “relative” minorizing coefficient $m(x, y) > 0$ such that $m(x, y)\gamma(x|y) \leq p(x)$.

It is important that $m(x, y)$ is symmetric, as it preserves the balance of flows from x to y and from y to x . Therefore $m(x, y)$ must be such that

$$m(x, y) \leq \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = L(x, y). \quad (25)$$

With the previously defined symmetric function $C(\cdot, \cdot)$ such that $C(\cdot, \cdot) \geq 1$, $m(x, y)$ can be written in the following form:

$$m(x, y) = \frac{1}{C(x, y)} \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\}. \quad (26)$$

Given $X_n = x \sim p(\cdot)$, the following “Markovian Minorizing” algorithm (MIR) may be used to generate $X_{n+1} \sim \pi(\cdot)$:

Algorithm MIR (Markovian Minorizing):

MI1. generate $y \sim \gamma(\cdot|x)$ and $r \sim U(0, 1)$

MI2. if $r \leq \alpha_{MI}(x, y)$, output $X_{n+1} = y$

MI3. else, output $X_{n+1} = x$

where

$$\alpha_{MI}(x, y) = \frac{m(x, y)\gamma(x|y)}{p(x)} = \frac{1}{C(x, y)} \min \left\{ \frac{\gamma(x|y)}{p(x)} \frac{p(y)}{\gamma(y|x)}, 1 \right\} \leq 1. \quad (27)$$

The transition kernel of Algorithm MIR is $P_{MI}(y|x) = \alpha_{MI}(x, y)\gamma(y|x)$ for all $x, y \in E$ and $x \neq y$, which satisfies detailed balance with respect to $p(\cdot)$:

$$p(x)P_{MI}(y|x) = p(x) \frac{m(x, y)\gamma(x|y)}{p(x)} \gamma(y|x) = P_{MI}(x|y)p(y).$$

As previously noted, IMIR with a deficient (absolute) minorizing coefficient m generates variates $y \sim p(\cdot)$ within the region $\{z : m\gamma(z) \leq p(z)\}$. Similarly, the relative minorizing coefficient $m(x, y)$ in the form of (26) may be deficient as an absolute minorizing coefficient, but it was chosen so that both x and y are in the region $\{z : m(x, z)\gamma(z|x) \leq p(z)\}$.

4.3 MIR and HA

MIR and MAR are equivalent because the acceptance probability of MIR in Equation (27) is identical with that of MAR in Equation (20). MIR therefore is also equivalent to HA. In fact, for any $\alpha_{MI}(x, y)$ defined by $m(x, y)$ in MIR, we set

$$\begin{aligned} s(x, y) &= m(x, y) \left(\frac{\gamma(x|y)}{p(x)} + \frac{\gamma(y|x)}{p(y)} \right) = \frac{m(x, y)\gamma(x|y)}{p(x)} \left(1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)} \right) \\ &\leq 1 + \frac{p(x)}{\gamma(x|y)} \frac{\gamma(y|x)}{p(y)}. \end{aligned} \tag{28}$$

Substituting this form of $s(x, y)$ into Equation (1) we obtain the same acceptance probability as $\alpha_{MI}(x, y)$ in HA. Conversely, for any $s(x, y)$ that defines $\alpha_{HA}(x, y)$, we let $m(x, y) = m_s(x, y) \leq L(x, y)$ as defined in Equation (13). Then, similar to Equation (14), Equation (27) yields $\alpha_{MI}(x, y) = \alpha_{HA}(x, y)$.

We have derived HA as MIR. Equations (13) and (28) show that there is a one-to-one mapping between the set of all symmetric functions $s(x, y)$ satisfying Hastings' condition (1) and the set of all symmetric functions $m(x, y)$ satisfying condition (25). However, $m(\cdot, \cdot)$ has a very intuitive interpretation of being the relative minorizing coefficients.

4.4 MIR and Algorithm M

Replacing $k(x, y)$ with $m(x, y) \leq L(x, y)$ in Equation (10), we obtain $\alpha_M(x, y) = \alpha_{MI}(x, y)$. MIR therefore is a special case of Algorithm M in which $k(x, y)$ is low enough to be a relative minorizing coefficient. The reverse is also true; that is, for every $k(x, y) > 0$, we define

$$m_k(x, y) = k(x, y) \min \left\{ \frac{p(x)}{k(x, y)\gamma(x|y)}, 1 \right\} \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\},$$

which is a relative minorizing coefficient because it satisfies the inequality (25):

Case 1. When $k(x, y) \geq H(x, y)$:

$$\begin{aligned}
m_k(x, y) &= \frac{1}{k(x, y)} \frac{p(x)}{\gamma(x|y)} \frac{p(y)}{\gamma(y|x)} \\
&\leq \frac{1}{\max \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\}} \frac{p(x)}{\gamma(x|y)} \frac{p(y)}{\gamma(y|x)} \\
&= \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = L(x, y)
\end{aligned} \tag{29}$$

Case 2. When $L(x, y) < k(x, y) < H(x, y)$:

$$\begin{aligned}
m_k(x, y) &= \begin{cases} \frac{p(x)}{\gamma(x|y)} & \text{if } \frac{p(x)}{\gamma(x|y)} \leq k(x, y) \leq \frac{p(y)}{\gamma(y|x)} \\ \frac{p(y)}{\gamma(y|x)} & \text{if } \frac{p(y)}{\gamma(y|x)} \leq k(x, y) \leq \frac{p(x)}{\gamma(x|y)} \end{cases} \\
&= \min \left\{ \frac{p(x)}{\gamma(x|y)}, \frac{p(y)}{\gamma(y|x)} \right\} = L(x, y)
\end{aligned} \tag{30}$$

Case 3. When $k(x, y) \leq L(x, y)$:

$$m_k(x, y) = k(x, y) \leq L(x, y)$$

Letting $m(x, y) = m_k(x, y)$ in Equation (27) yields:

$$\begin{aligned}
\alpha_{MI}(x, y) &= \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} \min \left\{ \frac{p(x)}{k(x, y)\gamma(x|y)}, 1 \right\} \frac{k(x, y)\gamma(x|y)}{p(x)} \\
&= \min \left\{ \frac{p(y)}{k(x, y)\gamma(y|x)}, 1 \right\} \min \left\{ \frac{k(x, y)\gamma(x|y)}{p(x)}, 1 \right\} = \alpha_M(x, y).
\end{aligned}$$

Hence Algorithm M is also a special case of MIR. They are equivalent.

5 Summary

We now summarize the relationship between Algorithm M , MAR and MIR by explaining what happens when $k(x, y)$ reduces from a very high value to a very low one.

Before doing so, we write Algorithm M in a two-stage form: Given $X_n = x \sim \pi(\cdot)$, then $X_{n+1} \sim \pi(\cdot)$ can be generated by

Algorithm L:

- L1. generate $y \sim \gamma(\cdot|x)$ and $r_1 \sim U(0, 1)$
- L2. if $r_1 > \min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\}$, output $X_{n+1} = x$ (MIR, type- x duplication)
- L3. else,
 - L3a. generate $r_2 \sim U(0, 1)$
 - L3b. if $r_2 > \min \left\{ \frac{p(y)}{k(x,y)\gamma(y|x)}, 1 \right\}$, output $X_{n+1} = x$ (MAR, type- y duplication)
 - L3c. else output $X_{n+1} = y$
- L4. endif

This allows us to classify the duplication of x either as a “type- x ” duplication, which occurs in Step L2, or as a “type- y ” duplication, which occurs in Step L3b. (The conditions in Steps L2 and L3b may be switched.) The probability of a type- x duplication is $1 - \min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\}$ and the probability of a type- y duplication is $\min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\} \left(1 - \min \left\{ \frac{p(y)}{k(x,y)\gamma(y|x)}, 1 \right\} \right)$. Thus the probability of x being duplicated is

$$\begin{aligned}
 & 1 - \min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\} + \min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\} \left(1 - \min \left\{ \frac{p(y)}{k(x,y)\gamma(y|x)}, 1 \right\} \right) \\
 &= 1 - \min \left\{ \frac{p(y)}{k(x,y)\gamma(y|x)}, 1 \right\} \min \left\{ \frac{k(x,y)\gamma(x|y)}{p(x)}, 1 \right\},
 \end{aligned}$$

which is the same as the probability of duplicating x in Algorithm M .

Case 1. When $k(x, y) \geq H(x, y)$: We start with a very high value of $k(x, y)$ such that $k(x, y) \geq H(x, y)$. Then $k(x, y)$ is a relative majorizing coefficient $M(x, y)$ and Algorithm M is MAR, utilizing only type- y duplications. There is a corresponding MIR with a relative minorizing coefficient $m_k(x, y)$ as defined in Equation (29), utilizing only type- x duplications. As $k(x, y) = M(x, y)$ decreases, both $\alpha_M(x, y) = \alpha_{MA}(x, y)$ and $m_k(x, y)$ increase. When $k(x, y) = M(x, y)$ decreases to $H(x, y)$, $m_k(x, y)$ increases to $L(x, y)$ and the acceptance probability $\alpha_M(x, y) = \alpha_{MA}(x, y)$ reaches its maximum value $\alpha_{MH}(x, y)$.

Case 2. When $L(x, y) < k(x, y) < H(x, y)$: When $k(x, y)$ further decreases below $H(x, y)$, it becomes “too deficient” for MAR to generate variates from $p(\cdot)$ with type- y duplications alone; type- x duplications are also needed to make Algorithm M equivalent to MH. As the value of $k(x, y)$ decreases further, we see fewer type- y duplications and more type- x duplications, but the acceptance probability $\alpha_M(x, y)$ remains at its maximum value $\alpha_{MH}(x, y)$. In this case, regardless of the value of $k(x, y)$, there is a corresponding relative majorizing coefficient $M_k(x, y) = H(x, y)$ as in Equation (22) and a corresponding relative minorizing coefficient $m_k(x, y) = L(x, y)$ as in Equation (30).

Case 3. When $k(x, y) \leq L(x, y)$: Further decreasing $k(x, y)$ below $L(x, y)$, we see Algorithm M becomes MIR, utilizing only type- x duplications, with $k(x, y)$ as a relative minorizing coefficient $m(x, y)$. There is a corresponding MAR with a relative majorizing coefficient $M_k(x, y)$ defined in Equation (23), utilizing only type- y duplications. As $k(x, y) = m(x, y)$ decreases from $L(x, y)$, $M_k(x, y)$ increases from $H(x, y)$, and the acceptance probability $\alpha_M(x, y) = \alpha_{MI}(x, y)$ decreases from its maximum value $\alpha_{MH}(x, y)$.

Algorithm M is a combination of MAR (which is HA), MIR (which is also HA) and MH (which is the optimal case of HA). It is not more general than HA, but it is easier to understand intuitively.

References

1. Barker, A. A. (1965) “Monte Carlo Calculations of the Radial Distribution Functions for a Proton-electron Plasma,” *Australian Journal of Physics*, 18, 119-33.
2. Billera, L. J. and Diaconis, P. (2001) “A Geometric Interpretation of the Metropolis-Hastings Algorithm,” *Statistical Science*, 16, 335-339.
3. Chib, S. and Greenberg, E. (1995) “Understanding the Metropolis-Hastings Algorithm,” *The American Statistician*, 49, 327-335.
4. Dongarra, J.; Sullivan, F. (2000) “Guest Editors’ Introduction: The Top 10 Algorithms,” *Computing in Science and Engineering*, 2, 22-23.

5. Hastings, W. K. (1970) “Monte Carlo Sampling Methods using Markov Chains and their Applications,” *Biometrika*, 57, 97–109.
6. Liu, J. S. (2001) *Monte Carlo Strategies in Scientific Computing*, New York: Springer-Verlag.
7. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller A. and Teller H. (1953) “Equation of State Calculations by Fast Computing Machines,” *The Journal of Chemical Physics*, 21, 1087-1092.
8. Minh, D. L. (2001) *Applied Probability Models*, Pacific Grove, CA: Duxbury Press.
9. Minh, D. L.; Minh, D. D. L. and Nguyen A. (2012) “Regenerative Markov Chain Monte Carlo for Any Distribution,” *Communications in Statistics - Simulation and Computation*, 41, 1745-1760.
10. Peskun, P. H. (1973). “Optimum Monte Carlo Sampling using Markov chains,” *Biometrika*, 60, 607–612.
11. Tierney, L. (1994) “Markov Chains for Exploring Posterior Distributions,” *The Annals of Statistics*, 22, 1701–1728.
12. von Neumann (1951) “Various Techniques Used in Connection with Random Digits,” *National Bureau of Standards*, Applied Mathematics Series, 12, 36-38.